

歩容特徴に基づく歩行者画像の新規生成を用いた歩容認証

吉野 弘毅^{1,a)} 中嶋 一斗^{1,b)} 安 正鎬^{1,c)} 岩下 友美^{2,d)} 倉爪 亮^{1,e)}

概要

歩容認証は、歩行者映像から対象者の歩き方の特徴を抽出して個人を識別する。歩行者映像には衣服等の人物に固有でない要素（共変量）も含まれるため、歩容認証では共変量の除去が重要である。共変量の除去には学習データが多様な共変量を含むことが望ましいが、プライバシーの問題や十分な歩行時間の確保が必要なことから、歩容認証のデータセットの大規模化は難しい。そこで本研究では、歩容と共変量の特徴を分離・抽出し、他の歩行者と共変量の特徴を入れ替えることによって歩行者画像を新規生成する。実験の結果、提案手法は高品質な歩行者画像を新規生成し、生成画像によるデータ拡張が識別学習の精度を向上させることを確認した。

1. はじめに

歩く姿である歩容は、個人によって固有である生体情報の一つである。歩容に基づく個人識別手法である歩容認証は、監視カメラ等の映像から歩容特徴を抽出することで歩行者を特定する。歩行者映像は遠隔から取得可能であり、歩行以外の被認証者の特別な動作を必要としないことから、歩容認証は大規模な犯罪捜査への応用が期待されている。歩行者映像には歩容だけではなく、個人を特定しない衣服や持ち物などの情報（共変量）が含まれており、歩容認証では共変量の除去と歩容特徴の抽出が重要である。

多くの従来手法 [1], [2], [3], [7] では、前処理によって歩行者映像からシルエットや骨格情報を抽出して背景などの共変量を除去した上で、歩容特徴を抽出してきた。前処理以外の手法としては、Disentanglement Representation Learning（分離表現学習）が挙げられる。分離表現学習は属性ラベルを教示する必要なく解釈可能な特徴表現を学習できる手法であり、分離表現学習を用いて歩容や共変量の特徴を入力から直接的に抽出する手法 [6], [9] が提案されて

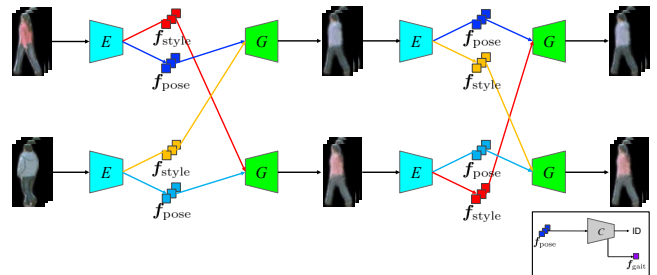


図 1 提案するネットワークの概要図。姿勢特徴 f_{pose} とスタイル特徴 f_{style} を分離するエンコーダ E 、姿勢特徴とスタイル特徴から歩行者画像を生成する生成器 G 、入力された画像の真贋判定を行う識別器 D 、姿勢特徴量をもとに人物 ID を識別する分類器 C から構成される

いる。しかし、分離表現学習は一般に教師なし学習であるため、分離性能は学習データの量とバリエーションに大きく依存する。また歩行者映像は、顔などのプライバシーに関する情報を多く含む上に、同じ脚を踏み出すまでの歩行周期が複数回含まれることが望ましいため、大規模なデータセットを作成するのは他の生体認証に比べて困難である。

そこで本研究では、分離表現学習によって分離した特徴を用いて歩行者画像を新規生成し、学習データを拡張する歩容認証手法を提案する（図 1）。提案手法では、データセット中の人数・歩行設定の種類・設定ごとの映像数に応じて、学習データのバリエーションおよび量を増やすことができるため、分離性能および特徴抽出器の汎化性能の向上が期待できる。具体的には、まず入力された歩行者映像は、歩容に依存する特徴（姿勢特徴）および外観に関する共変量の特徴（スタイル特徴）に分離される。次に、異なる人物の二つの歩容画像間で、姿勢特徴を交換して歩行者画像を新規生成し、学習データに追加する。最後に、時系列の姿勢特徴群から抽出される歩容特徴を用いて識別を行う。実験では、提案手法によって生成された歩行者画像に対する定性的に評価し、識別精度を算出することで、提案手法による生成画像のデータ拡張に対する有用性について検証を行う。

2. 提案手法

提案手法で用いるネットワークの概要図を図 1 に示す。ここで、姿勢特徴は、時系列の歩容画像列において、人物

¹ 九州大学

² NASA / Caltech ジェット推進研究所

a) yoshino@irvs.ait.kyushu-u.ac.jp

b) k_nakashima@ait.kyushu-u.ac.jp

c) ahn@irvs.ait.kyushu-u.ac.jp

d) yumi.iwashita@jpl.nasa.gov

e) kurazume@ait.kyushu-u.ac.jp

に依存する時変性を持つ特徴（全身形状や関節角度など）であり、スタイル特徴は、人物に依存しない時不変性を持つ特徴（服装・歩行方向など）と定義する。提案手法におけるベースラインとして採用したのは、本手法と同じくカラー画像を入力とする分離表現学習を用いた手法である Zhang らの手法 [9] である。

2.1 特徴の分離・抽出

入力された歩容映像は、エンコーダによって姿勢特徴およびスタイル特徴に分離・抽出される。Zhang ら [9] 同様、歩容映像の特徴が適切に分離・抽出されるように、同一人物で歩行条件の異なる二つの歩容画像からそれぞれ得られた姿勢特徴から、次の姿勢特徴類似性損失を計算する。

$$\mathcal{L}_{\text{pose-sim}} = \left\| \frac{1}{n_1} \sum_{t=1}^{n_1} \mathbf{f}_{\text{pose}}^{(c_1,t)} - \frac{1}{n_2} \sum_{t=1}^{n_2} \mathbf{f}_{\text{pose}}^{(c_2,t)} \right\|_2^2 \quad (1)$$

2.2 歩行者画像の生成

次に、エンコーダ E によって抽出された姿勢およびスタイル特徴を、生成器 G に入力して歩容画像を生成する。生成画像および元画像は識別器 D に入力され、入力画像が生成画像か実画像であるかを判別される。

Zhang ら [9] に従い、同一の歩容映像中からランダムに取り出した時刻の異なる二つのフレームから、スタイル特徴と姿勢特徴をそれぞれ分離・抽出する。時刻 k のフレームから抽出されたスタイル特徴および時刻 l のフレームから抽出された姿勢特徴を生成器 G に入力して生成した歩容画像、および姿勢特徴の抽出元となった時刻 l のフレームから、次の再構成損失を計算する。

$$\mathcal{L}_{\text{recon}} = \sum_{c \in \{c_1, c_2\}} \sum_{\substack{k, l \in \{1, \dots, n\} \\ k \neq l}} \left\| G(\mathbf{f}_{\text{style}}^{(c,k)}, \mathbf{f}_{\text{pose}}^{(c,l)}) - I^{(c,l)} \right\|_1 \quad (2)$$

実画像と生成画像ともに識別器 D に入力し、次の敵対的損失を計算する。ただし、敵対的損失の定義には、予備実験の結果から、最も学習が安定していた RaLSGAN [5] を採用した。

$$\mathcal{L}_{\text{adv}} = \sum_{i,j \in \{c_1, c_2\}} \sum_{t=1}^n \left(\left(D(I^{(j,t)}) - \mathbb{E} \left[D \left(G(\mathbf{f}_{\text{style}}^{(i,t)}, \mathbf{f}_{\text{pose}}^{(j,t)}) \right) \right] - 1 \right)^2 + \left(D \left(G(\mathbf{f}_{\text{style}}^{(i,t)}, \mathbf{f}_{\text{pose}}^{(j,t)}) \right) - \mathbb{E} \left[D(I^{(j,t)}) \right] + 1 \right)^2 \right) \quad (3)$$

2.3 生成画像の再エンコード

生成された画像も識別学習に活用するために、生成画像は再度エンコーダ E に入力され、2.1 節と同様に姿勢特徴とスタイル特徴に分離される。分離された特徴の元となった画像の特徴を、生成画像に正確に保持させるために、再度エンコードする前後での両特徴の差分から、エンコーダ E の一貫性を保証する損失を計算する。

$$\mathcal{L}_{\text{consis}}^{\text{style}} = \sum_{i,j \in \{c_1, c_2\}} \sum_{t=1}^n \left\| \mathbf{f}_{\text{style}}^{(i,t)} - E_{\text{style}}(G(\mathbf{f}_{\text{style}}^{(i,t)}, \mathbf{f}_{\text{pose}}^{(j,t)})) \right\|_1 \quad (4)$$

$$\mathcal{L}_{\text{consis}}^{\text{pose}} = \sum_{i,j \in \{c_1, c_2\}} \sum_{t=1}^n \left\| \mathbf{f}_{\text{pose}}^{(j,t)} - E_{\text{pose}}(G(\mathbf{f}_{\text{style}}^{(i,t)}, \mathbf{f}_{\text{pose}}^{(j,t)})) \right\|_1 \quad (5)$$

2.4 元画像の循環再構成

さらに、再エンコードした特徴から元画像を再構成する。異なる二つの歩容画像 I^A, I^B に対して、エンコード、特徴交換生成、再エンコードの順で処理することで、 I^A の姿勢特徴 $E_{\text{pose}}(G(\mathbf{f}_{\text{style}}^B, \mathbf{f}_{\text{pose}}^A))$ 、 I^A のスタイル特徴 $E_{\text{style}}(G(\mathbf{f}_{\text{style}}^A, \mathbf{f}_{\text{pose}}^B))$ が得られる。これら二つの特徴をもとに、次の循環再構成損失を定義する。

$$\mathcal{L}_{\text{cycle}} = \sum_{\substack{a,b \in \{A,B\} \\ a \neq b}} \left\| G(E_{\text{style}}(G(\mathbf{f}_{\text{style}}^a, \mathbf{f}_{\text{pose}}^b)), E_{\text{pose}}(G(\mathbf{f}_{\text{style}}^b, \mathbf{f}_{\text{pose}}^a))) - I^a \right\|_1 \quad (6)$$

この損失は、Zhu ら [10] が報告するように教師データのない画像生成の高品質化への貢献が期待され、提案手法における特徴の分離性能の向上にも寄与すると考えられる。

2.5 姿勢特徴による人物識別

すべての抽出された姿勢特徴は分類器 C に入力され、被験者 ID の確率分布が出力される。Zhang ら [9] に従い、分類器 C によって出力された人物 ID の確率分布から、次の交差エントロピーを計算する。

$$\mathcal{L}_{\text{id}} = \sum_{c \in \{c_1, c_2\}} \left(\frac{1}{\sum_{t=1}^n \omega_t} \sum_{t=1}^n -\omega_t \mathbf{y}^T \log \left(C(\mathbf{f}_{\text{pose}}^{(c,1)}, \dots, \mathbf{f}_{\text{pose}}^{(c,t)}) \right) \right) \quad (7)$$

学習では、以上 7 つの損失に対して重みをかけて足し合わせて同時に最適化する。

3. 実験

実験では、CASIA-B [8] データセットを用いて、生成画像品質の定量的評価および定性的評価、識別精度の定量評価を行った。CASIA-B は、124 名から構成されるデータセットで、前半の 74 名を学習に、後半の 50 名を評価に用いた。また、CASIA-B には NM (標準)・BG (バッグを所持)・CL (NM・BG とは異なる服装) の三つの設定が含まれている。本実験で用いたベースラインは、Zhang [9] らの論文を元に筆者らが実装したものである。Adaptive Instance Normalization (AdaIN) [4] とは、スタイル変換タスクにおいて広く用いられる、スタイル変換の自由度を向上させる正規化処理であり、ベースラインからの改善点の一つとして提案手法に追加されている。

3.1 生成画像品質の定性評価

二つの歩行者画像間で、姿勢特徴を交換して生成した画像を図 2 に示す。1, 2 行目は、それぞれ生成画像における姿勢特徴、スタイル特徴の元となった画像である。そのた

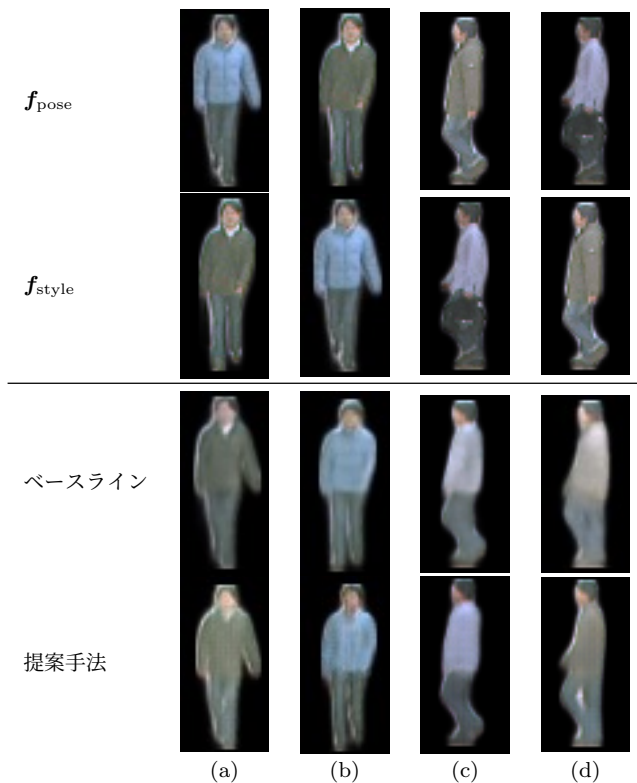


図 2 各手法における姿勢特徴交換による生成画像. 1 行目: 姿勢特徴の元画像. 2 行目: スタイル特徴の元画像.

め、分離表現学習による分離性能が高い場合、1 行目の姿勢をしていながら 2 行目の服装をしている画像が生成されると期待される。ベースラインの生成画像を見ると、(a), (b) は服装を忠実に再現できているものの、(c), (d) は服装の色が薄くなってしまっており元画像の再現に失敗している。それに対して提案手法は、すべての生成画像において対応する元画像の服装および姿勢を再現していることがわかる。

3.2 識別精度の定量評価

ベースラインと提案手法における CASIA-B に対する Rank-1 認証率の値を表 1 に示す。表 1 より、すべての設定において提案手法は識別精度を向上させていることがわかる。しかし CL では、提案手法含めて全手法で他の設定よりも精度が 60%程度小さくなっている。ベースラインの論文 [9] 中では本実験の値よりも 30%以上高く報告されているため、ベースラインの再実装の再現度が不十分であることが原因の一つであると考えられる。

表 1 CASIA-B [8] における識別精度 [%]

手法	NM	BG	CL
ベースライン	90.2	86.1	24.3
提案手法	93.9	89.4	26.8

4. まとめと今後の予定

本研究では、分離表現学習の分離性能を向上させるために、入力画像から抽出した特徴を用いて歩行者画像を新規生成して、学習データを拡張する新しい歩容認証手法を提案した。提案手法では、まずカラーの歩容画像から姿勢とスタイル特徴を直接的に分離・抽出する。分離した姿勢特徴を、異なる人物の映像間で交換することによって歩行者画像を新規生成し、学習データに加える。実験の結果、提案手法は生成画像の品質と分離表現学習における分離性能を向上させ、提案手法によって生成された合成画像は、データ拡張に有効であることが示された。

今後は、提案手法によって生成した画像の定量的評価を行っていく予定である。

謝辞

本研究の一部は、JST 科学技術イノベーション創出に向けた大学フェローシップ創設事業 JPMJFS2132 および JSPS 科研費 JP20H00230 の助成を受けたものである。

参考文献

- [1] Chao, H., He, Y., Zhang, J. and Feng, J.: Gaitset: Regarding gait as a set for cross-view gait recognition, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, No. 01, pp. 8126–8133 (2019).
- [2] Fan, C., Peng, Y., Cao, C., Liu, X., Hou, S., Chi, J., Huang, Y., Li, Q. and He, Z.: Gaitpart: Temporal part-based model for gait recognition, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14225–14233 (2020).
- [3] Feng, Y., Li, Y. and Luo, J.: Learning effective gait features using LSTM, *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR)*, pp. 325–330 (2016).
- [4] Huang, X. and Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1501–1510 (2017).
- [5] Jolicoeur-Martineau, A.: The relativistic discriminator: a key element missing from standard GAN, *arXiv preprint arXiv:1807.00734* (2018).
- [6] Li, X., Makihara, Y., Xu, C., Yagi, Y. and Ren, M.: Gait Recognition via Semi-supervised Disentangled Representation Learning to Identity and Covariate Features, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13309–13319 (2020).
- [7] Liao, R., Yu, S., An, W. and Huang, Y.: A model-based gait recognition method with body pose and human prior knowledge, *Pattern Recognition*, Vol. 98, p. 107069 (2020).
- [8] Yu, S., Tan, D. and Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition, *Proceedings of the IEEE Inter-*

- national Conference on Pattern Recognition (ICPR)*, Vol. 4, pp. 441–444 (2006).
- [9] Zhang, Z., Tran, L., Yin, X., Atoum, Y., Liu, X., Wan, J. and Wang, N.: Gait recognition via disentangled representation learning, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4710–4719 (2019).
- [10] Zhu, J.-Y., Park, T., Isola, P. and Efros, A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks, *Proceedings of the IEEE international conference on computer vision (ICCV)*, pp. 2223–2232 (2017).